Supporting Information

Evolutionary history of the p53 family DNA-binding domain: insights from an *Alvinella pompejana* homolog

Qiang Zhang¹, Dimitrios-Ilias Balourdas^{2,3}, Bruno Baron⁴, Alon Senitzki⁵, Tali E. Haran^{*,5}, Klas G. Wiman^{*,6}, Thierry Soussi^{*,7,8,9} & Andreas C. Joerger^{*,2,3,9}

¹ Department of Neuroscience, Biomedicum, Karolinska Institutet, Stockholm, Sweden

² Institute of Pharmaceutical Chemistry, Goethe University, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany

³ Buchmann Institute for Molecular Life Sciences and Structural Genomics Consortium (SGC), Maxvon-Laue-Str. 15, 60438 Frankfurt am Main, Germany

⁴ Plateforme de Biophysique Moléculaire, Centre de Ressources et de Recherches Technologique (C2RT), Institut Pasteur, 75015 Paris, France

⁵ Department of Biology, Technion-Israel Institute of Technology, Technion City, Haifa 32000, Israel

⁶ Department of Oncology-Pathology, Bioclinicum, Karolinska Institutet, Stockholm, Sweden

⁷Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden

⁸ Sorbonne Université, UPMC Univ Paris 06, 75005 Paris, France

*correspondence:

joerger@pharmchem.uni-frankfurt.de (A.C.J), thierry.soussi@sorbonne-universite.fr (T.S.), bitali@technion.ac.il (T.E.H.), or klas.wiman@ki.se (K.G.W.).

⁹These authors contributed equally

EST_N72937	1	SSS	APS	ΡΥΙ	DN	ΤQ	AL :	SΡ	S S I	_ H M	ΝНΙ	Q	5 P /	ΑP	ΤV	ΡS	ΝT	ΡY	's c	ΞE Υ	ΥG	FΕ	IS	5 F (знα	2 S	КЕТ	54
GO114003.1	1	SSS	APS	ΡΥΙ	P D N	ΤQ	AL	SΡ	SSI		N H I	LQS	5 P /	A Ρ	TVI	ΡS	ΝT	ΡΥ	' S (5 E '	ΥG	FΕ	IS	5 F (ЭНС	S S	КЕТ	54
GO114002.1	1							ΗН	HP :	5 1	N H I	LQS	5 P A	A X	ТΥ	ΡS	ΝT	ΡY	' S (5 E '	ΥG	ΧE	IS	5 F (зно	2 S	КЕТ	39
Protein crystallized	1												9	5 M	ΤVΙ	ΡS	ΝT	ΡY	' S (5 E '	Y G	FΕ	IS	; F (Ο Η C	2 S	КЕТ	27
EST_N72937	55	кѕт	т w т	FSE	E S L	КК	LF	VR	M A ⁻	гтα	۲P	V R F	۶КП	ΓV	HQ	ΡP	ΑG	S٧	11	RAI	МР	ΙY	νĸ	(P '	ЕH	VQ	EVV	108
GO114003.1	55	ΚSΤ	т w т	FSE	E S L	КΚ	LF۱	VR	M A ⁻	гтс	CР	V R F	⁼ К 1	Г٧	HQI	ΡΡ	ΑG	S۷	11	RAI	ИP	ΙΥ	ΝK	(P	ΕH	VQ	EVV	108
GO114002.1	40	кѕт	т w т	FSE	ΞSL	КΚ	LF	VR	M A ⁻	гтс	P١	V R F	• к 1	гν	HQI	ΡΡ	ΑG	sν	11	RAI	ΜP	ΙY	٧ĸ	(P)	ЕΗ	VQ	ΕVV	93
Protein crystallized	28	КЅТ	т w т	FSE	E S L	КК	LF	VR	MA	ГΤΟ	EΡ	V R F	⁻ К 1	ΓV	HQI	ΡP	ΑG	S۷	11	RAI	МР	ΙY	νĸ	(P	ΕH	VQ	EVV	81
EST_N72937	109	KRC	ΡΝΗ	IAT	ГКЕ	ΗN	EDI	ΗP	AP.	ГΗЦ	. V I	RCE	HK	< L	AS	ΥV	ΕD	ΡY	ΥT	5 R (Q S	νı	ΙP	'Q	ΕH	ΡQ	AGA	162
GO114003.1	109	KRC	ΡΝΗ	IAT	ГКЕ	ΗN	ΕDI	ΗP	AP ⁻	ГΗЦ	. V I	RCE	ΕHK	< L	AS	ΥV	ΕD	ΡY	ΥC	5 R (Q S	VΙ	ΙP	'Q	ΕH	ΡQ	AGA	162
GO114002.1	94	KRC	ΡΝΗ	ΙΑΤ	ГКЕ	ΗN	EDI	ΗP	AP.	ГΗЦ	. V I	RCE	ΕHK	< L	AS	ΥV	ΕD	ΡY	ΥT	G R (Q S	VΙ	I P	' Q '	ΕH	ΡQ	AGA	147
Protein crystallized	82	KRC	ΡΝΗ	IAT	ГКЕ	ΗN	EDI	ΗP	A P ⁻	ГΗЦ	. V I	r c e	ΕHM	< L	AS	ΥV	ΕD	ΡY	ΥC	G R (Q S	VΙ	IP	'Q	ΕH	ΡQ	AG A	135
EST_N72937	163	EWV	ΤNL	YQI	- M C	FS	s c '	VG	GLI	NRF	R P	IQ۱	/ I F	τ	LEI	ΗE	Gν	νL	. G I	RQ /	ΑV	ΕV	RΙ	C	A C	ΡG	RDR	216
GO114003.1	163	ΕWV	ΤNL	YQI	- M C	FS	sс	VG	GLΙ	NRF	R Ρ	IQ۱	/ I F	τ	LEI	ΗE	G۷	VL	GI	RQ	ΑV	ΕV	RΙ	C	A C	ΡG	RDR	216
GO114002.1	148	ΕWV	ΤNL	YQI	- M C	FS	sс	VG	GΙΙ	NRF	RΡ	IQ۱	/ I F	Т	LEI	ΗE	G۷	VL	G I	RQ	ΑV	ΕV	RΙ	C	A C	ΡG	RDR	201
Protein crystallized	136	EWV	ΤNL	YQF	- м с	FS	sс	VG	GLI	NRF	RΡ	١Q١	/ I F	τ	LEI	ΗE	Gν	νL	. G I	RQ.	A V	ΕV	RI	C	A C	ΡG	RDR	189
							_																					
EST_N72937	217	RAE	ЕТА	AD	P N K	QQ	KR	ΡA	NKI	N T I	IS -	ΓEΛ	AT S	5 V (GΡ	ΑL	ΚP											251
GO114003.1	217	RAE	ЕТА	ADI	P N K	Q -																						228
GO114002.1	202	RAE	ЕТА	AD	NK	QQ	KR	ΡA	NKI	N T I	IS T	ΓEΛ	AT S	5 V (GΡ	ΑL	ΚP											236
Protein crystallized	190	RAE	ЕТА	ADI	P N K	QQ	ΚR																					204

Figure S1. Translation of *Alvinella pompejana* p53 homolog ESTs and sequence of the protein construct used for X-ray crystallography.



Figure S2. Thermal unfolding of annelid worm p53 family DBDs monitored by DSF with SYPRO orange. Normalized raw data of DSF analysis at a heating rate of 180 °C/h for the DBDs of *Alvinella pompejana* and *Capitella teleta* (with and without N-terminal hexahistidine tag) showing increased thermostability of the *Alvinella pompejana* DBD. In both cases, the presence of an N-terminal hexahistidine tag had a minor effect on DBD stability, lowering the T_m by about 1 °C.



В

Variants	AP substitutions	Other substitutions
p.Q144R	12	54
p.V157I	30	718
p.Q167E	10	24
p.M169V	13	32
p.R174K	13	61
p.H178N	10	110
p.G187D	15	56
p.1232N	29	109
p.M246L	10	357
p.1255F	103	128
p.D259H	12	204
p.F270V	29	205
p.V274I	12	319
p.P309S	20	8

Figure S3. Distribution of p53 cancer variants in the DNA-binding domain of the human p53 protein. (A) For each position in the protein (p53 residues 100-300, ref NP_000537.3) the number of variations is shown. AP substitutions (red, lower part): number of p53 cancer variants leading to an amino acid substitution corresponding to the wild-type residue in the *Alvinella pompejana* protein. Other substitutions (blue, upper part): number of p53 variants leading to an amino acid substitution to the wild-type residue in the *A. pompejana* protein. (B) Table showing the most frequent p53 variants identified in the above graph. The frequency of the different cancer mutations is based on the 2017 issue of the TP53 database (http://p53.fr, 80,000 variants) [1].

		LSH motif cysteine cluster
		· · · · · · · · · · · · · · · · · · ·
		124 125 141
Homo saniens n53/P04637	110	
Homo saniens p63/09H3D4	197	
Homo saniens p73/015350	127	
Mus musculus p53/0549C9	116	
	107	
Mus_musculus_posposo	120	
Xenonus tropicalis p53/E7A9/10	02	
Xenopus tropicalis p63/E67GN7	03	
Xenopus tropicalis p73/E6TKT0	130	
Danio rerio p53/G1K2L5	88	
Danio_rerio_p53/A7YY17	01	
Danio_rerio_p73/080177	144	
Callorhinchus milii n53/G9/11/8	100	
Callorhinchus milii p63/G9/11.9	210	
Callorhinchus milii p73/G9/1M0	142	AKSATWTYSPIIKK - IYCOLAKTCPIOAKVTS 172
Lethenteron japonicum p53/A0A01/4D4F9	102	AKSATWTYSPDINK-LYCOLGKTCPVOLRVAT 132
Lethenteron japonicum p63/A0A0U3KDC1	183	AKSATWTESTKI KK - I YCOMSKI CPAFIRTST 213
Lethenteron japonicum p73/A0A0U4B546	94	AKSATWTYSPDIKK - I YCOLAKTCPLOEKVIS 124
Ciona intestinalis/04H300	129	PKSAOYTYSPIINK-IEVKMNVTCPIKEKCAR 159
Ciona intestinalis/Q4H278	161	PKSAPETYSYSIOK - IEVKMNENCPIKERCSP 191
Branchiostoma floridaelC3XPU2	188	AKSATWTYSEVINK - I YVRMAITCPVKEKTRI 218
Branchiostoma floridaelC3YXH3	119	SKKLAWTWSEPCKM-LYADENKLCPENEKTVR 149
Ixodes scapularis	79	R N S T S W T Y S N A K R K - L Y T N M N F F C S V H F L T F R 109
Aedes aegyptilO171M1	163	SGKSSWMESSBINK - VEVKMGOACTENISYOA 193
Aedes aegyptilO171M5	168	S D K T S WMF S B K L O K - L F V K MG N T C T F N V S O L P 198
Daphnia pulexIE9FW11	67	TKATPWIYSTKLKK - LEVSMDKACPISERMDN 97
Drosophila melanogaster/09N6D8	100	PPKSLWMYSIPLNK-LYIRMNKAENVDVOEKS 130
Caenorhabditis elegans Q20646	238	S S DMAFAISSEHEKYLWTKMGCLVPIOVKWKL 269
Mya arenaria Q9NGC8	159	TKSTTWTYSDILKK-LYVRMATTCPVRFKTLR 189
Mytilus edulis Q1AMZ8	212	TKSTTWTYSESLKK - LYVRMATTCPIRFKCLR 242
Lottia_gigantea V4A869	158	TKSTTWTYSEMLKK-LYVRMATTCPVRFKTCH 188
Crassostrea_gigas K1RC48	356	T K S T T W T Y S E S L K K - L Y V R M A T T <mark>C</mark> P V R F K S Q R 386
Octopus_bimaculoides A0A0L8G262	160	TKSTTWTYSEKLDK-LYVRMATT <mark>C</mark> PVRFKTLQ 190
Alvinella_pompejana	53	TKSTTWTFSESLKK-LFVRMATTCPVRFKTVH 83
Capitella_teleta R7UHV7	150	TKSTTWTYSEPIRK-LFVRMATTCPVRFKTDL 180
Helobdella_robusta T1EZJ4	1	MKTP <mark>C</mark> PFLFRTSR 13
Nematostella_vectensis A8DPD6	131	PKSAPWTYSHQLQK-LL <mark>C</mark> RMKCLVPVRLVFRS 161
Nematostella_vectensis	109	VANPDWIYSTSQNK-LYIKPQTP <mark>C</mark> PMKFSVTG 139
Trichoplax_adhaerens B3RZS6	151	TKATNNTYSTMLKK-LFIKMECLFPIHITIER 181
Amphimedon_queenslandica A0A1X7UD97	188	PKTVPFTYSNLMKR - AYIKRDSTVGMTFSFSK 218
Amphimedon_queenslandica A0A1X7URI3	59	DGSRKFIQTADKEKHLIIKKDVNFALSFSSDT 90
Monosiga_brevicollis A9UZX3	118	ARAIGWTYSPILNT - LFTPMDYS <mark>C</mark> PIRFATNE 148

Vertebrates

Invertebrates

Figure S4. Conservation of the three-cysteine cluster in the loop-sheet-helix motif in the DNA-binding domain of p53 family proteins. Sequences were aligned using Jalview [2] with manual editing based on known crystal structures. UniProt accession codes are given after the species and name of the protein. This alignment suggests a gradual appearance of the three cysteines during p53 evolution, with Cys141 (human p53 numbering) possibly already present in the last common ancestor of all extant animals, Cys135 appearing shortly before the radiation of vertebrates and Cys124 last, as a vertebrate p53 specific variation, first observed in bony fishes.

						• •	•						•	•						•	•	•	•	•	•	•	•	• •	
A. pompejana homolog	31	VP-	S N T P	Y <mark>S</mark> G	E Y G I	F <mark>E</mark> I	S F	QHC	Q S K	ΕT	<mark>к</mark> s т	ти	V <mark>T</mark> F	S E	S L	ккι	. F \	/ <mark>R</mark> //	A T	T C F	v <mark>R</mark>	FΚ	T V J	H <mark>Q</mark> F	PP	۱G <mark>S</mark>	٧I	RAMP	96
C. teleta homolog R7UHV7	128	- P W	/ <mark>R</mark> HTP	Y <mark>S </mark> G	E F G I	F <mark>S</mark> I	S F	QQC	Q S K	ΕT	кsт	TW	V <mark>T</mark> Y	S E	ΡΙ	R K L	. F \	/ <mark>R</mark> //	A T	T C F	י v <mark>R</mark>	FΚ	T D I	l <mark>q</mark> f	PF	IG A	۷I	RAMP	193
H. sapiens p53 P04637	97	VP-	S Q <mark>K</mark> T	YQG	S Y G I	F <mark>R</mark> L	GΕ	LH ·	- <mark>S</mark> G	ΤА	ĸsν	и <mark>т</mark> с	T Y	S P	ΑL	N K N	ΛFC	C <mark>Q</mark> L	ΑK	T C F	v v <mark>Q</mark>	LW	v D/	S T F	PF	' G T	RV	RAMA	161
H. sapiens p63 Q9H3D4	165	IP-	SNTC	YPG	р н <mark>s</mark> I	- Dv	S F	QQ -	s s	ΤА	K S A	л т и	V T Y	sт	ΕL	ккι	. Y (Q I	ΑK	T C F	י ו <mark>ע</mark>	ιк	vм	ТРР	P P C	G A	V I	RAMP	229
G. gallus p53 P10360	13	VP-	STED	YGG	DFDI	F <mark>R</mark> V	GF	V E -	AG	ΤА	ĸsν	T C	T Y	S P	VL	N K \	/ Y (C <mark>R</mark> L	ΑK	ΡCF	v v <mark>Q</mark>	V R	۷G۱	VAF	PF	' G <mark>S</mark>	S L	RAVA	77
		••		_	••	•••		_				•	•	.	•••	••			•	•	_		•	•	• • •	• •	_	-	
A. pompejana homolog	97	IYV	'	IVQE	V V <mark>K</mark>	R C P	' <mark>N</mark> H	ATT	r k e	ΗN	E D -	- H P	P A P	T H	LV	R C E	H	< L -	A S	YVE	DP	YT	G <mark>R (</mark>	Q S \	/ []	i P <mark>Q</mark>	EH	P <mark>Q</mark> A G	161
C. teleta homolog R7UHV7	194	IYN	1 K P E H	IVQE	V V T	R C P	' <mark>N</mark> H	A T T	r k e	ΗN	EN	- НР	P A P	' <mark>к</mark> н	LV	r c e	E H I	< L -	AQ	YKC	DDH	ΥT	L R 🤇	Q S \	V V	ΙРΗ	E P	P <mark>Q</mark> A G	258
H. sapiens p53 P04637	162	I Y <mark>K</mark>	QSQH	IMTE	V V <mark>R</mark>	<mark>R</mark> C P	<u>н</u> н	ERC	C <mark>S D</mark>) - S	DG	L A	A P P	<mark>Q</mark> H	LΙ	RVE	G	NL R	VΕ	Y L 🕻	D D R	ΝT	FR	H <mark>S</mark> \	/ V \	/ P Y	E P	P <mark>E</mark> VG	226
H. sapiens p63 Q9H3D4	230	V Y <mark>K</mark>	КАЕН	IV <mark>TE</mark>	V V <mark>K</mark>	<mark>R</mark> C P	' <mark>N</mark> H	ELS	RE	FΝ	E G C		A P P	S H	LΙ	RVE	G <mark>I</mark>	NS F	I A <mark>Q</mark>	YVE	DP	ΤT	G R 🤇	QS	/ L \	/ P Y	ΕP	P <mark>Q</mark> V G	296
G. gallus p53 P10360	78	V Y K	: к <mark>ѕ</mark> е н	IVA E	V V R	R C P	нн	ERC	GGG	GТ	DG	LA	A P A	Q H	LΙ	r v e	G	N P Q	A R	ΥΗ	DE	ΤТ	K R I	н <mark>s</mark> \	/ v \	/ P Y	ΕP	P <mark>E</mark> V G	143
			_	•_•	••	•		•••		•	•••	•••	•				•		•	•	••	•	•_			_		_	
A. pompejana homolog	162	AEV	V V <mark>T N</mark> L	. Y <mark>Q</mark> F	M C F	s s c	ΣVG	GL	N R R	R P I	QV	I F T	T L E	HE	- G	VVI	LG	R Q A	V E	V R	ICA	СР	G R	DRI	RAI	E E T	A A	D P N K	227
C. teleta homolog R7UHV7	259	A E V	V V T N L	. F <mark>Q</mark> F	M C F	s s c	ΣVG	GL	N R R	R P I	QV	I F <mark>1</mark>	T L E	HD	- G	R V I	LG	R <mark>Q</mark> A	V E	VR	ICA	СР	GR	DRI	R A I	DE <mark>S</mark>	A A	Q P K T	324
H. sapiens p53 P04637	227	S D C	C <mark>TT</mark> IH	HY <mark>N</mark> Y	M C N	s s c	MG	GM	N R R	R P I	LΤ	111	T L E	DS	<mark>S</mark> G	NLI	LG	R <mark>N</mark> S	FE	۷R	V C A	СР	GR	DRI	RTI	E E E	NL	R K <mark>K </mark> G	293
H. sapiens p63 Q9H3D4	297	ΤEF	T T V L	. Y <mark>N</mark> F	M C N	s s c	ΖVG	GM	N R R	ΡI	LI	ו ע ו	T L E	TR	DG	Q V I	LG	RRC	FE	A R	ICA	СР	GR	DRI	K A I	DED	S I	R K <mark>Q Q</mark>	363
G. gallus p53 P10360	144	S D C	C T T V L	. Y <mark>N</mark> F	M C N	s s c	MG	GM	N R R	ΡI	LΤ	I L <mark>1</mark>	T L E	GΡ	GG	Q L I	LG	R R C	FE	۷R	V C A	СР	GR	DRI	кі	ΕEE	N F	RKRG	210

Figure S5. Sequence alignment and CvP-bias of p53 family DBDs. Sequences were aligned using Jalview [2]. Accession codes are given after the species and the name of the protein. CvP-bias = $(N_{charged} - N_{polar})/\text{length x 100}$. $N_{charged} =$ number of charged amino acids (E, D, K, R; highlighted in green), $N_{polar} =$ number of polar amino acids (S, T, N, Q; highlighted in orange). CvP values were calculated using the domain boundaries shown in the alignment. Higher CvP values have been associated with higher thermostability of orthologs [3]. This correlation is not seen when comparing the p53 family DBDs of the thermophilic *A. pompejana* and the mesophilic *C. teleta* annelids, which have CvP values of 0.5 and 1.5, respectively. These values were slightly higher than for the human p63 DBD (-1.5) and the human p53 DBD (-1.0). The CvP value of the chicken (*G. gallus*) p53 DBD was much higher (9.1), supporting the notion that the high stability of chicken p53 compared with the human homolog is, at least in part, due to the higher number of charged surface residues [4]. Buried side chains are marked with a dot. Unmarked residues have side chains with a solvent accessibility of $\geq 10\%$ in the crystal structure of the human p53 DBD (PDB entry 2XWR, chain A) as calculated with the software NACCESS, which uses a probe radius of 1.4 Å for calculating solvent-accessible areas.



Figure S6. EMSA of p53 family DBDs binding to Con1 DNA at various temperatures. DBDs (nominal concentration of 2400 nM) were incubated with Con1 DNA for 60 min at the indicated temperature and then run on a 6% polyacrylamide gel (37.5:1 (acrylamide: bisacrylamide ratio) at 4 °C.



Figure S7. Polyacrylamide gel electrophoresis of representative DBD samples used in the present study. The following abbreviations are used: Hs p53, human p53; Hs p.R175H, human p53 hot spot mutant R175H (structural mutant); Hs p.R273H, human p53 hot spot mutant R273H (DNA-contact mutant); Xl p53: *Xenopus laevis* p53; Om p53, *Oncorhynchus mykiss* (rainbow trout) p53; Nc p53, *Notothenia coriiceps* (Antarctic fish) p53 (not used in the present study); Gg p53, *Gallus gallus* (chicken) p53; Ap p53, *Alvinella pompejana* p53 homolog.

	Alvinella pompejana p53 homolog
	DNA-binding domain
Data Collection	
Space Group	<i>P</i> 1
<i>a</i> , b, c (Å),	55.90, 69.99, 114.85
α, β, γ (°)	88.07, 89.92, 81.51
Molecules/asymmetric unit	8
Resolution (Å) ^a	49.7-1.92 (1.95-1.92)
Unique reflections	121,535
Completeness (%) ^a	92.5 (93.0)
Multiplicity ^a	3.7 (3.8)
$R_{ m merge}$ (%) ^a	5.6 (83.5)
CC(1/2) ^a	0.998 (0.630)
Mean $I/\sigma(I)^a$	12.5 (1.7)
Refinement	
<i>R</i> _{work} , (%) ^b	18.4
$R_{\rm free}$, (%) ^b	22.5
No. of atoms	
Protein ^c	11,898
Zinc	8
Water	860
Ethylene glycol	16
RMSD bonds (Å)	0.007
RMSD angles (°)	0.83
Mean B (Å ²)	38.2
Ramachandran favored (%) ^d	97.7
Ramachandran outliers (%) ^d	0.1
PDB entry	7PC6

Table S1. X-ray data collection and refinement statistics

^aValues in parentheses are for the highest-resolution shell.

 ${}^{b}R_{\text{work}}$ and $R_{\text{free}} = \sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum |F_{\text{obs}}|$, where R_{free} was calculated with 5 % of the reflections chosen at random and not used in the refinement.

°Number includes alternative conformations.

^dMolProbity [5] statistics

SI references

- 1. Leroy, B. et al. The TP53 website: an integrative resource centre for the TP53 mutation database and TP53 mutant analysis. *Nucleic Acids Res* **41**, D962-9 (2013).
- 2. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. & Barton, G.J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-91 (2009).
- 3. Holder, T. et al. Deep transcriptome-sequencing and proteome analysis of the hydrothermal vent annelid Alvinella pompejana identifies the CvP-bias as a robust measure of eukaryotic thermostability. *Biol Direct* **8**, 2 (2013).
- 4. Khoo, K.H., Andreeva, A. & Fersht, A.R. Adaptive evolution of p53 thermodynamic stability. *J Mol Biol* **393**, 161-75 (2009).
- 5. Williams, C.J. et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293-315 (2018).